Title: NH EPHT Program's use of Semantic Metadata
Author: Tom Lambert
Reviewers: Matt Cahillane, Chia-Hui Chawla
Last Review Date: September 26, 2012

Forward: background information can be found in the many useful articles returned by searching for 'semantic metadata' on the web.

Introduction

Use of semantics will probably drive the next major revision of how the web and web searches operate. It allows for machines to understand more of the meaning in a query term and draw results based on that understanding.

Using semantics involves assigning metadata to content which has syntax tagged with standardized notation and including reference to the preferred ontology.

Ontologies are at the heart of the semantic web. Ontologies give objects meaning and place them into a conceptual model. Conceptual models and ontologies can be created by anyone for any reason. Some models are published openly for use by others and some are private for use within an organization.

Example 1: Semantic Web for Earth and Environmental Terminology (SWEET)

An example of a potentially useful conceptual model and ontology is the Semantic Web for Earth and Environmental Terminology (SWEET; http://sweet.jpl.nasa.gov/; Appendix A), which has been developed by NASA. It is highly suggested that the reader go to the website above and download the protégé application which helps navigate the SWEET ontology. SWEET bases its conceptual model on its wordly understanding of things in the universe. NASA's SWEET conceptual model is depicted in Figure 1.

Figure 1. The SWEET conceptual model.

If EPHT finds NASA's conceptual model and ontologies useful, it could begin tagging its NCDM metadata (which refers to a piece of EPHT Content) with reference to the SWEET ontology. For example, NCDM Metadata for air pollution includes the term National Ambient Air Quality Standard (NAAQS).

If the metadata tagged NAAQS with the SWEET, ontology the NCDM content would be placed within the context of SWEET. Figure 2 shows some of this context. Therefore, a semantic enabled web search engine (such as Google is starting to develop) will understand that the NAAQS are part of an environmental control strategy which is both part of a governing body product and a control strategy, which are both the products of a decision as a human activity.

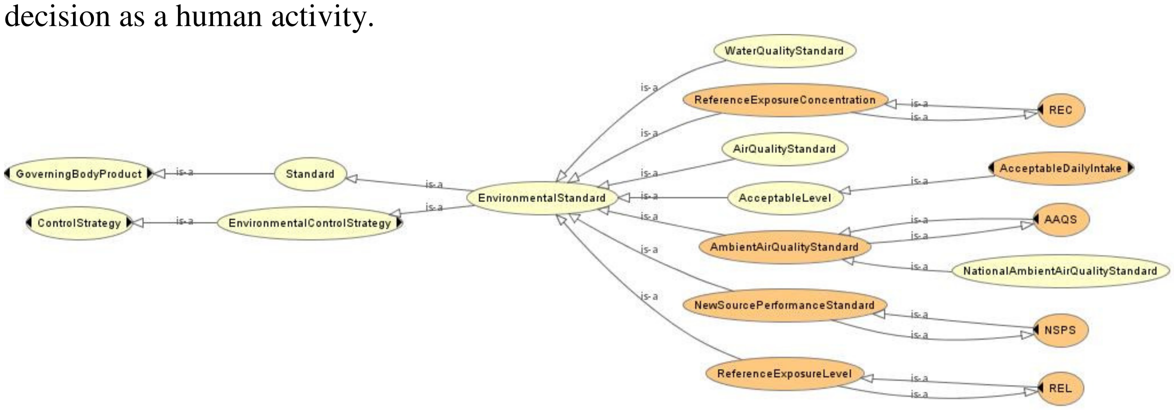

Figure 2. SWEET's model of a mid level class of 'Environmental Standard'.

Taking this one logical step forward – if you searched Google for AAQS, the search results might show you that within AAQS there are National AAQS. Perhaps there would also be World AAQS (which aren't included in the ontology) and so on.

As the ontology develops, users (such as EPHT and its partners) can create subclasses (as needed) of the NAAQS for each chemical species and/or aerodynamic particulate size for which there is a standard. A subclass might be ozone which links to chemical, trace gas, toxic substance, etc. in the SWEET ontology.

There are an enormous number of possible connections in something like the SWEET conceptual schema. There are too many for one individual to create and manage. The good news is that the semantic web calls for an extensible system which can begin simply where entities and relationships can be defined as needed. The power of extensibility means when a conceptual model is developed communally, and one conceptual model can link at any level to any level of a separate model developed by different organizations who are experts in their respective area, the model becomes extremely rich.

For example, the SWEET schema contains some basic periodic table information about compounds and elements but it does not contain detailed information about their bonds, behaviors, reactivity, etc. This is because SWEET's developers may not be experts in this area or they just don't need that information. However, that doesn't mean they can't link compounds and elements to CombeChem RDF conceptual schema (http://semanticweb.com/semantic-chemistry_b10684) to fill in the details.

Taking another step forward – semantic metadata for web objects other than text can also be created. For instance, a data set, graph, map, table, picture, etc. can be attributed semantic metadata. Further, a value in a table, an axis on a graph, or a legend in a map could all have semantic metadata which link the object to its ontology.

Example 2: CDC's Content Management Ontological Approach for EPHT
CDC staff at the NEPHT are working on developing an ontology to provide a logical and coherent foundation for Tracking Program, to make system components more substitutable, and to better position Tracking Program to handle expanding content areas. CDC's ontology is focused heavily on the placement of things within EPHTs operational view and represents generalized concepts and relationships for organizing environmental health information, defining related measures, and implementing business rules governing the analysis and visualization of those measures (Figure 3).,

CDC content management ontology supports development of a CMS designed to facilitate a multi-dimensional information hierarchy, algorithmically generated reports and visualizations, and a user interface for more efficient content creation. This content ontological approach lays the foundation for future work to incorporate extant Tracking content metadata and leverage existing ontology and vocabularies.

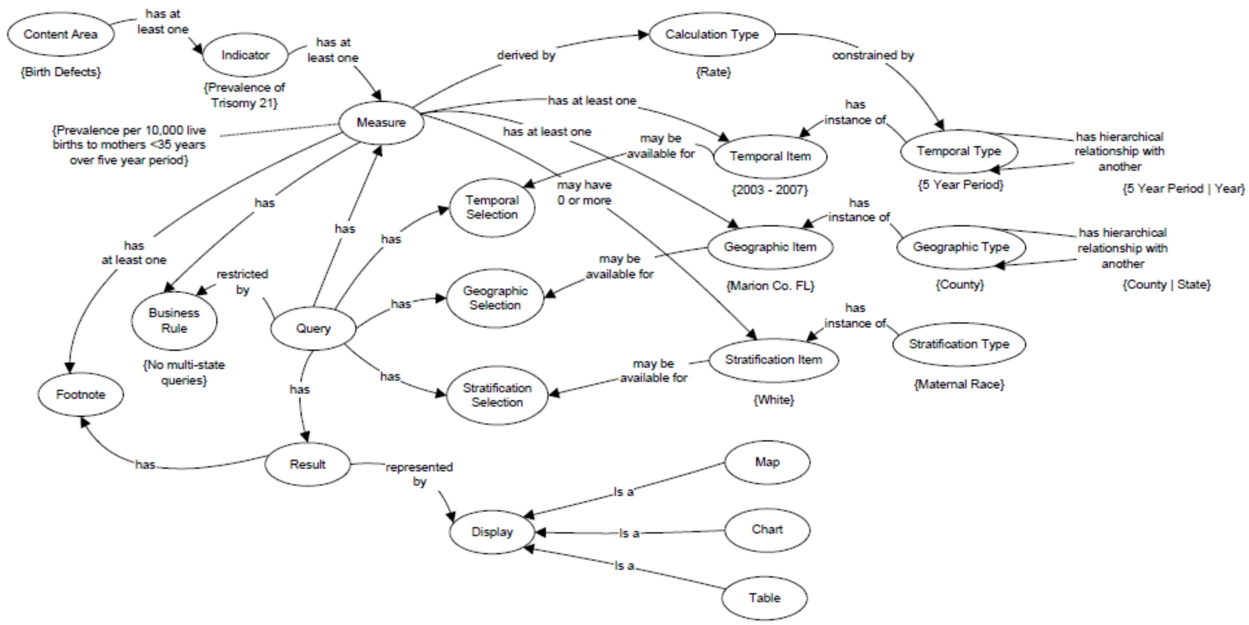

Figure 3. CDC Content Management Ontology

The approach will facilitate data linkage, increase content access through improved search capabilities, and move Tracking closer to its goals for building a sustainable network and advancing science and research. Figure 4 shows an example of CMS future state as development of the CMS ontology continues, additional entities and relationships concerning metadata and existing standards will be considered for inclusion. This CMS framework may be a reference for NH EPHT to establish semantic metadata structure.

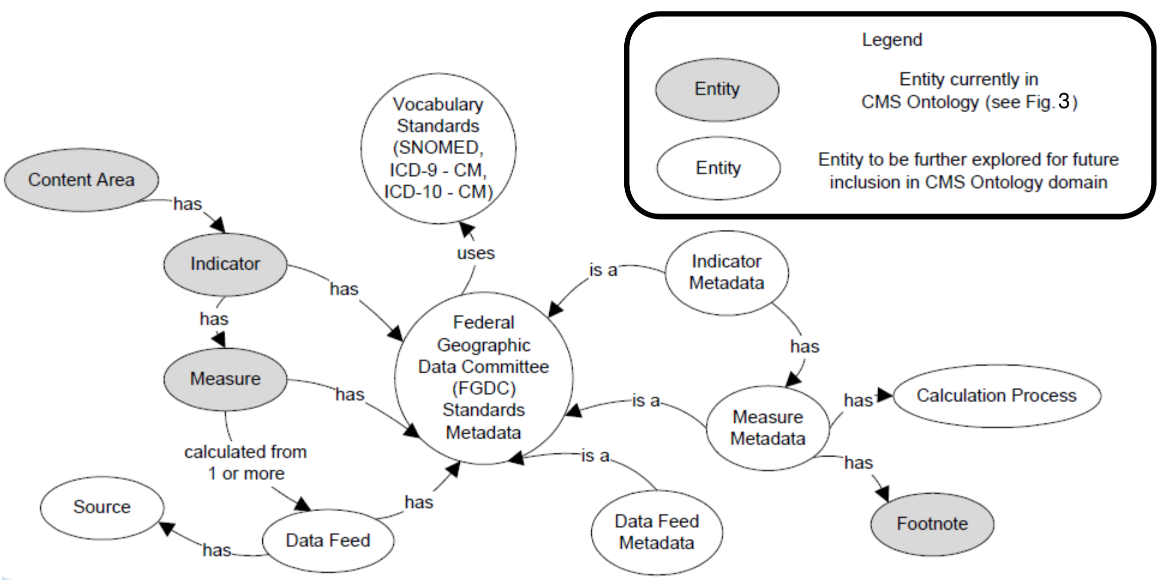

Figure 4. CDC Content Management System Future State

Comparison of SWEET and CDC ontologies

While the diagram of ovals and arrows are similar in appearance, the two ontologies couldn't be more dissimilar in their approach. CDC's approach to developing its ontology is different that SWEET. It focuses more on the practical aspect of things within its practical operational view. For instance, CDC focuses on a measure as a 'thing' while SWEET (by my observation) would call the same thing a representation of a property, which can apply to anything which has a property. CDC is able to flatten its ontological structure by narrowing its scope to focus on its own business needs. SWEET's scope is universal and because of this many classes may be needed to relate one object to another. To summarize, SWEET has developed a universal ontology that CDC's ontology could fit into. However, to fit one into the other could result in a significant amount of work which needs to be performed before a meaningful, valid, and useful product is created.

Using semantic metadata in NH EPHT

NH EPHT is participating with the NH Health Web-based Interactive System for Direction and Outcome Measures (WISDOM) Project. WISDOM's goals are to

1) improve accessibility and utility of public health data sets
2) put all data into the context of performance management at all levels of the organization
3) help people publish integrated data to the web

The WISDOM project is working to develop web content which represents the various aspects of our business and data. We have created a conceptual model of how the NH Division of Public Health Services plans, organizes, performs, and measures its work in order to help it improve on how decisions are made regarding its business (Figure 5). So far the conceptual model has been represented using an Entity – Relationship diagram (E-R diagram) and in a relational database format. We will generate metadata for each piece of content, entity within the conceptual model, and data set. The conceptual model was not built with the semantic web in mind, but it is serving as a syntactic foundation to link our business to external semantic ontologies, such as SWEET. Take for example one cross sectional view of our conceptual model taken from the standpoint of 'all work is performed through an activity':

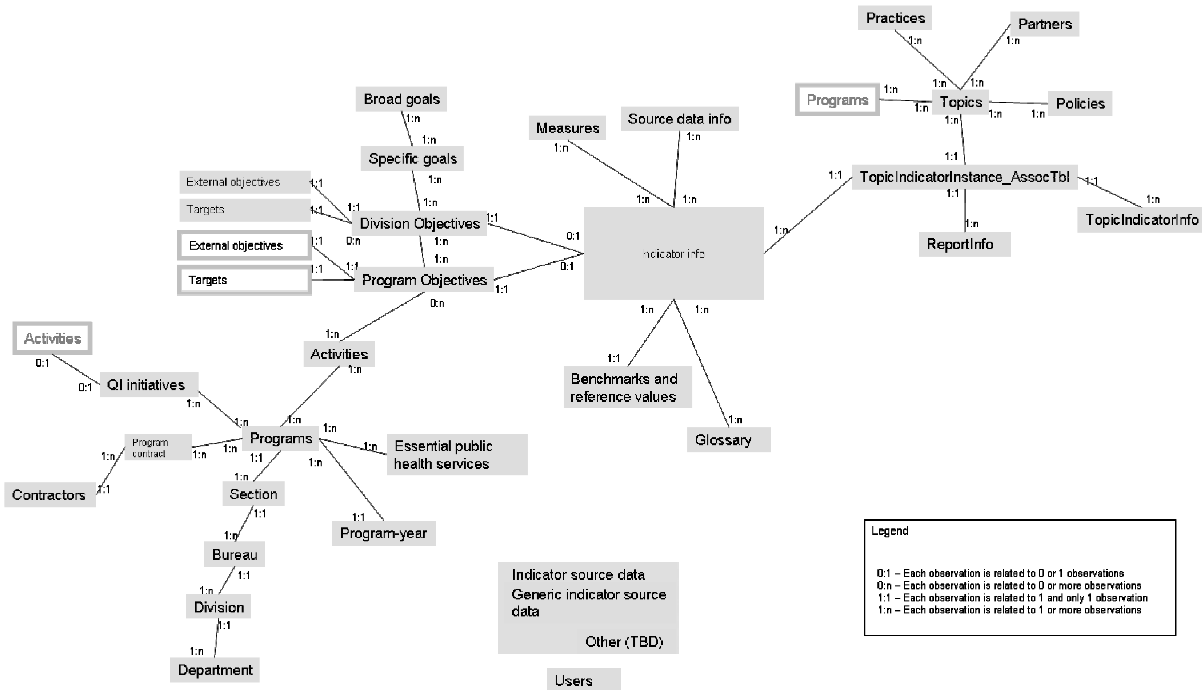

Figure 5: WISDOM Conceptual schema

The WISDOM project views our business as such (this is not exhaustive):

1) People working in public health perform activities
2) People are grouped strategically into Programs, Sections, and Bureaus based on their knowledge skills and ability.
3) Grouping people helps reduce friction and assist in managing activities
4) Activities take resources to perform
5) Resources are finite so activities are performed strategically
6) All activities are intended to directly or indirectly improve the health of a population
7) People decide what activities to do based on the identification of the health problem, available resources, and our knowledge, skill, and ability
8) Activities are supposed to be based on evidence of previously successful methods if available
9) Activities either improve the health of a population or they do not (within a given time period)
10) The magnitude of health improvements can usually be measured
11) Data exist to represent the health improvement for some health problems
12) Indicators define the process of measuring whether an activity is improving the health of the population
13) The performance of an activity can be based on the magnitude of the health improvement
14) Activities can be changed, added, or removed by decision makers based on performance

15) Decision makers need activities to have pre set measurable objectives which indicate the expected magnitude of the improvement by a given date and/or whether the activity is going to succeed or has succeeded
16) Several indicators may be needed in order to measure whether an activity or series of activities are succeeding
17) Unpredicted external events can impact the effectiveness of the activity

Looking into this cross section we begin to see how entities in our conceptual model fit into SWEET's ontology. For example:

1) Our activities are the result of human decisions and human decisions are a sub class of SWEET's Human Activities top-level.
2) People are living things within the Matter top-level. People who perform activities and who are the beneficiary of activities are both sub classes of people
3) Data is captured within SWEET's representation top-level
4) Whether an activity succeeded or not is a binary property of the activity outcome
5) Indicators are representations of a mathematical process
6) Etc.

If we go a step further and actually link our syntax with SWEET's ontology and adding classes where needed and tag syntax in the content metadata using applicable standards, we will have successfully placed it within SWEET's ontology. Note - SWEET's developers may not want to include certain of our entities in their conceptual model, but perhaps other ontologies exist which are more appropriate.

Availability and gaps in ontologies related to EPHT

Many of the most important entities within EPHT (by extension of WISDOM) intersect with existing ontologies.

1) CDC and NEPHT's ontology and CMS is most syntactically similar to NH EPHT's needs because our business needs overlap. However, these ontologies do not fit into a coherent universal view which raises concerns of their extensibilities.
2) NASA's SWEET ontology covers a lot of the high level needs of NH EPHT regarding semantics of the physical and chemical environment. NH's conceptual schema fits more in with SWEET than with CDC and NEPHT.
3) The geonames (http://www.geonames.org/) can cover much of the geospatial semantics.
4) CombeChem RDF covers a lot of the chemical, reactivity, and toxicity ontology
5) Health outcome related ontologies may be covered by Health Level - Seven (HL7) or other ontologies (though this was not evaluated in detail).

These additional areas need to be evaluated for available ontologies: policy and law, people and their attributes, society, politics, finances and currency.

Ontology-specific syntax must be tagged correctly within our metadata. Using multiple ontologies could mean that multiple tagging processes will be implemented.

Future work

NH EPHT may use pieces of the above listed ontologies to help refine its business in the near term and we will work to ensure our content management system is structured as closely as possible to their most logical and extensible aspects.

NH EPHT CDC and NEPHT ontologies to help refine its own business but it can't use them because in some aspects it would be a step backwards. For instance, we view indicators as an attribute of a process. It's a definition of how and why something is to be measured and not as a physical object which has FGDC metadata (for us, FGDC metadata is metadata which is conformed to a information management strategy).

NH EPHT views the SWEET ontology as being developed more in line with how we are trying to conceptualize our business. NH EPHT views the SWEET ontology as being more extensible than CDC/NEPHT ontologies.

NH EPHT may propose a project to do two things

1) Explore development of content metadata with tagged syntax which uses third party ontology, such as SWEET and CDC, to return context via a web service which we wouldn't otherwise have.
2) Explore development of content metadata and our own simple ontology so that search results are placed into the full context of our conceptual model.
3) Collaborate with the NEPHT SND Metadata workgroup to share information on the use of semantic metadata.

Reference

Semantic Web for Earth and Environmental Terminology (SWEET)
http://sweet.jpl.nasa.gov/

CombeChem RDF conceptual schema
http://semanticweb.com/semantic-chemistry_b10684

The geonames
http://www.geonames.org/

National EPHT Standards and Network Development (SND) Metadata Usability Team.
https://ephtn.sharepointsite.net/snd/MD_UsabilityWiki/Home.aspx

Metadata Creation and Usability Whitepaper_A Primer for Content Creation, Searchability, and Messaging
https://ephtn.sharepointsite.net/snd/MD_UsabilityWiki/MD%20Messaging%20and%20Notes%20Whitepaper.aspx

James B. Jellison, MPH and Craig A. Kassinger, BS. Centers for Disease Control and Prevention; Atlanta, Georgia. *Expanding Environmental Public Health Tracking Through an Ontological Approach for Content Management*. Poster presented at Fall AMIA 2011
https://ephtn.sharepointsite.net/snd/NASG1/Presentations/Peer-to-Peer%20Presentations/CDC%27s%20Tracking%20Ontology%20Project/CDC%20Tracking%20Ontology%20AMIA%202011%20poster.pdf

Acknowledgements: This publication was supported by grant number #1U38EH000947-01 from the National Center for Environmental Health (NCEH) at the Centers for Disease Control and Prevention (CDC) in Atlanta. Its content is solely the responsibility of the authors and does not necessarily represent the official views of the CDC.